

# Обучение машинного перевода без параллельных текстов\*

Скиднов Е. А. Бахтеев<sup>1</sup> О. Ю. Стрижов<sup>2</sup> В. В.

<sup>1</sup>Московский физико-технический институт

<sup>2</sup>Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

Данная задача посвящена исследованию алгоритма обучения машинного перевода без параллельных предложений. Использование параллельных текстов для задачи машинного перевода требует слишком большой базы предложений всех переводимых языков, что является ресурсоемкой задачей для некоторых пар непохожих языков. Особенностью исследуемого алгоритма является то, что для перевода используется кодирование и декодирование текста во внутреннем представлении. Данный алгоритм использует 2 модели нейронной сети Seq2Seq для перевода с одного языка на другой и обратно. Цель данного исследования заключается в том, чтобы сделать вектора скрытых пространств этих двух моделей как можно более похожими. Для демонстрации работоспособности метода будет использован вычислительный эксперимент машинного перевода между двумя похожими языками: русским и украинским.

**Ключевые слова:** *машинный перевод, нейросеть, Seq2Seq.*

## 1 Введение

Решается задача выбора оптимального метода для машинного перевода. В классическом решении данной задачи требуется порядка нескольких миллионов параллельных предложений. Доступ к таким крупным базам данных есть далеко не у всех, и не для каждой языковой пары они существуют. Кроме того такое обучение затратит большое количество ресурсов, как производительных, так и временных.

Существует ряд подходов к построению оптимального метода обучения. Предлагается использовать рекуррентные нейронные сети с короткой и долгой памятью и нейронные сети, в которых реализовано внимание. В других методах используются нейронные сети, которые осуществляют перевод в два этапа. Такой метода называется Seq2Seq.

Данная работа посвящена последнему методу последовательного перевода. Предлагается с помощью первой рекуррентной нейронной сети, основанной на долгой памяти перевести входящую последовательность в вектор, а с помощью второй перевести этот вектор в выходную последовательность на нужном нам языке. Данный метод позволяет гораздо быстрее обучить нейронную сеть переводу с одного языка на другой, в связи с использованием ей предыдущего опыта и наличию у нее памяти и внимания.

Проверка и анализ метода проводятся с помощью алгоритма BLEU (Bilingual evaluation understudy) для проверки качества текста, переведенного с одного языка на другой на паре языков русский-украинский.

---

\*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А.В. Грабовой, О.Ю. Бахтеев, В.В. Стрижов, Eric Gaussier, координатор Малиновский Г.С. Консультант: Бахтеев О. Ю.